

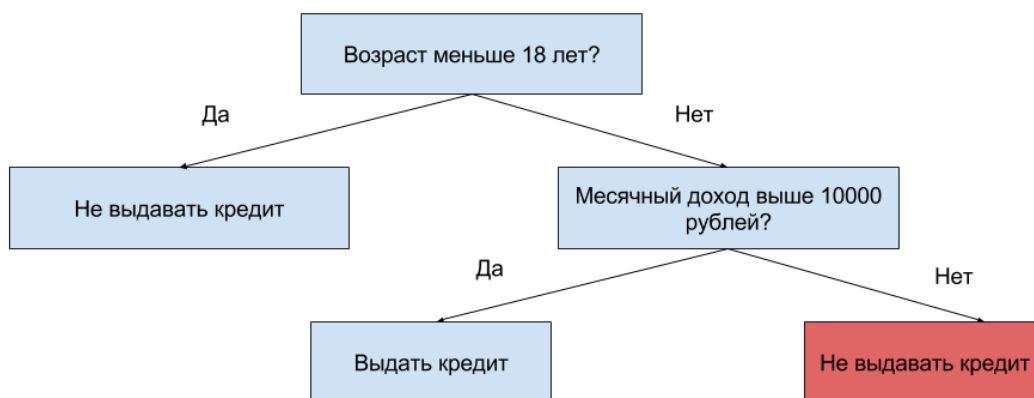
1. Какую проблему решают критерии останова и стрижка деревьев?

1 point

- ☐ Они позволяют строить маленькие деревья, и поэтому помогают ускорить процесс обучения.
- ☐ Они позволяют выбирать более подходящие условия для каждой вершины, и поэтому помогают восстанавливать более сложные закономерности.
- ☒ Они уменьшают сложность дерева, и поэтому помогают решать проблему переобучения.

2. Рассмотрим решающее дерево для решения задачи кредитного скоринга, которое основано на двух признаках: возрасте и месячном доходе.

1 point



Пусть дана следующая выборка из пяти объектов (первый признак — возраст, второй — месячный доход):

[20, 8000]

[15, 15000]

[28, 9500]

[24, 30000]

[30, 20000]

Для сколько из них будет принято положительное решение о выдаче кредита?

2

3.

1 point

Мы пытаемся найти лучшее разбиение вершины m , и хотим оценить качество конкретного способа, который разобьёт вершину m на вершины l и r . Обозначим через $|X_m|$, $|X_l|$, $|X_r|$ количество объектов в вершинах m , l и r соответственно, через $H(X)$ — значение критерия информативности на выборке X .

По какой из формул следует вычислять ошибку такого разбиения?

- ☐ $H(X_l) + H(X_r)$
- ☐ $\frac{|X_l|}{|X_m|} \frac{H(X_l)}{H(X_m)} + \frac{|X_r|}{|X_m|} \frac{H(X_r)}{H(X_m)}$
- ☒ $\frac{|X_l|}{|X_m|} H(X_l) + \frac{|X_r|}{|X_m|} H(X_r)$
- ☐ $\frac{H(X_l)}{H(X_m)} + \frac{H(X_r)}{H(X_m)}$

4. Мы всё ещё пытаемся найти лучшее разбиение для вершины m , и теперь хотим вычислить значение критерия Джини для левой вершины в конкретном разбиении. В выборке три класса, и распределение объектов между ними в левой вершине выглядит так: $p_1 = 0.9$, $p_2 = 0.07$, $p_3 = 0.03$. Чему же будет равно значение критерия Джини?

1 point

0.5

5. Можно ли решать задачу регрессии с помощью решающих деревьев?

1 point

- ☐ Нет, нельзя — деревья могут выдавать столько различных ответов, сколько в дереве листьев, то есть конечное число. А в задаче регрессии бесконечно много возможных ответов.
- ☐ Нет, нельзя — критерии информативности зависят от распределения объектов по классам, такие распределения нельзя построить в задачах регрессии.
- ☒ Да, можно — достаточно лишь выбрать критерий информативности, оценивающий разброс вещественных ответов.

6. Как можно использовать категориальные признаки в решающем дереве в подходе, который строит n -арные деревья?

1 point

- ☒ Вершина разбивается на столько дочерних вершин, сколько различных значений у категориального признака. Каждая из дочерних вершин соответствует одному значению.

- ☐ Вершина разбивается на столько дочерних вершин, сколько различных значений у категориального признака. Каждая из дочерних вершин соответствует значениям с определённой частотой встречаемости.
 - ☐ Вершина разбивается на n дочерних вершин, причём n — параметр метода. Каждая из дочерних вершин соответствует случайно выбранному подмножеству значений признака.
 - ☐ Вершина разбивается на n дочерних вершин, причём n — параметр метода. Каждая из дочерних вершин соответствует определённому подмножеству значений, которое выбирается на основе специального критерия.
7. Какую форму будет иметь разделяющая поверхность, построенная деревом с условиями вида $[x^j < t]$ в вершинах? Считайте, что в выборке два признака.
- 1 point
- ☒ Кусочно-линейная кривая. Каждый её участок — линейная функция, коэффициенты которой задаются в одном из листьев.
 - ☐ Кусочно-постоянная кривая, участки которой будут параллельны осям координат. Это связано с тем, что условие в каждой вершине разделяет объекты с помощью прямой, параллельной одной из осей координат.
 - ☐ Разделяющая поверхность может иметь совершенно любую форму.